

02 - A

Web Scrapping XPATH

Club de Programación Creativa

Por Alejandro Ramos

@arhcoder

01 Internet

Conceptos básicos

02 HTML

Estructura de sitios web

03 XPATH

Extrayendo de un HTML

04 Código

Aplicando scraping

03

XPATH

XPATH

Lenguaje que permite construir
expresiones para recorrer y
procesar **documentos XML**

XML Path Language

HTML es un lenguaje de programación creado como estándar para representar estructuras de datos.

RECORDANDO QUÉ ES HTML

HTML se escribe en archivos y estos archivos representan "documentos", tal cual como un PDF, XML, etc, son la estructura de información fija

```
18 <title>Tic, Tac, Toe!</title>
19 </head>
20
21 <body onload="reset()">
22   <div class="top-panel">
23     <div class="titular"><h1>¡TIC, TAC, TOE!</h1><
24     <div id="player-turn"><h2 id="turn-text">Inici
25   </div>
26   <div class="content">
27     <div class="box-empty" id="box1" onclick="
28     <div class="box-empty" id="box2" onclick="
29     <div class="box-empty" id="box3" onclick="
30     <div class="box-empty" id="box4" onclick="
31     <div class="box-empty" id="box5" onclick="
32     <div class="box-empty" id="box6" onclick="
33     <div class="box-empty" id="box7" onclick="
34     <div class="box-empty" id="box8" onclick="
35     <div class="box-empty" id="box9" onclick="
36   </div>
37   <div class="bottom-panel">
38     <div class="scores" id="scores">
39       <div id="score-x-box">
40         <h2 id="x-score">0</h2>
43         <img src="svg/o.svg" class="score-icor
```


HTML es un lenguaje de programación creado como estándar para construir **estructuras de sitios web**

HTML se escribe en archivos y estos archivos representan **"documentos"**, tal cual como un PDF, XML, etc, son la estructura de información fija

```
18 <title>Tic, Tac, Toe!</title>
19 </head>
20
21 <body onload="reset()">
22   <div class="top-panel">
23     <div class="titular"><h1>¡TIC, TAC, TOE!</h1><
24     <div id="player-turn"><h2 id="turn-text">Inici
25   </div>
26   <div class="content">
27     <div class="tictactoe">
28       <div class="box-empty" id="box1" onclick="
29       <div class="box-empty" id="box2" onclick="
30       <div class="box-empty" id="box3" onclick="
31       <div class="box-empty" id="box4" onclick="
32       <div class="box-empty" id="box5" onclick="
33       <div class="box-empty" id="box6" onclick="
34       <div class="box-empty" id="box7" onclick="
35       <div class="box-empty" id="box8" onclick="
36       <div class="box-empty" id="box9" onclick="
37     </div>
38   </div>
39   <div class="bottom-panel">
40     <div class="scores" id="scores">
41       <div id="score-x-box">
42         <h2 id="x-score">0</h2>
45         <img src="svg/o.svg" class="score-icor
```



```

<body>
  <header>
    <h1>Mi Página WEB</h1>
  </header>

  <section>
    <article>
      <h2>Sección 1</h2>
      <p>Este es un párrafo dentro de un artículo. También hay una imagen:</p>
      
      <button>Botoncito</button>
    </article>
  </section>

  <section>
    <article>
      <h2>Sección 2</h2>
      <p>Otro párrafo dentro de otro artículo. Aquí hay un enlace y un formulario:</p>
      <a href="https://www.google.com">Enlace a google.com</a>
      <form>
        <label for="nombre">Nombre:</label>
        <input type="text" id="nombre" name="nombre">
        <button type="submit">Enviar</button>
      </form>
    </article>
  </section>

  <footer>
    <p>&copy; 2024 Mi Página Web</p>
  </footer>
</body>

```

Mi Página WEB

Sección 1

Este es un párrafo dentro de un artículo. También hay una imagen:



Botoncito

Sección 2

Otro párrafo dentro de otro artículo. Aquí hay un enlace y un formulario:

[Enlace a google.com](https://www.google.com)

Nombre:

© 2024 Mi Página Web

```

<body>
  <header>
    <h1>Mi Página WEB</h1>
  </header>

  <section>
    <article>
      <h2>Sección 1</h2>
      <p>Este es un párrafo dentro de un artículo. También hay una imagen:</p>
      
      <button>Botoncito</button>
    </article>
  </section>

  <section>
    <article>
      <h2>Sección 2</h2>
      <p>Otro párrafo dentro de otro artículo. Aquí hay un enlace y un formulario:</p>
      <a href="https://www.google.com" data-bbox="100 550 350 580">Enlace a google.com</a>
      <form>
        <label for="nombre">Nombre:</label>
        <input type="text" id="nombre" name="nombre">
        <button type="submit">Enviar</button>
      </form>
    </article>
  </section>

  <footer>
    <p>© 2024 Mi Página Web</p>
  </footer>
</body>

```

XML es casi lo mismo que HTML

Mi Página WEB

Sección 1

Este es un párrafo dentro de un artículo. También hay una imagen:



Sección 2

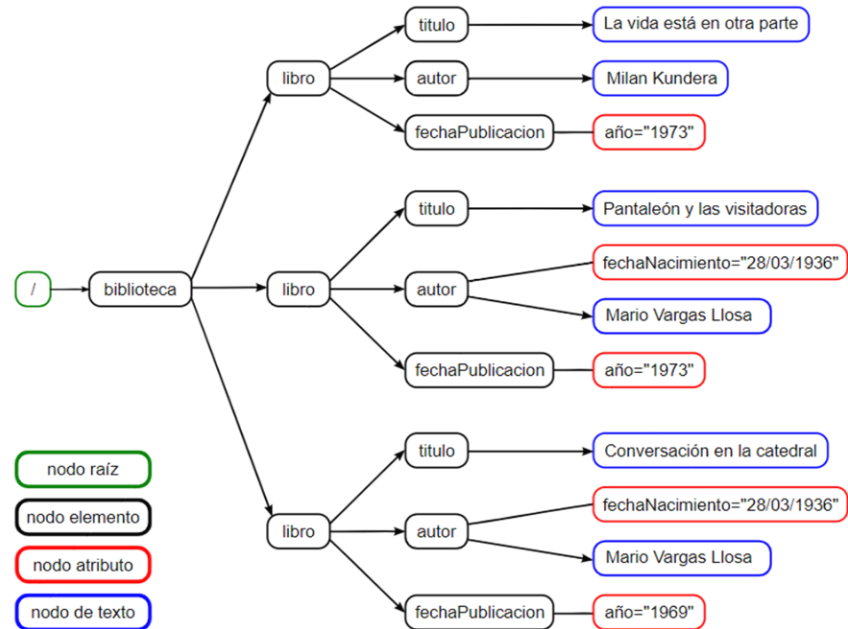
Otro párrafo dentro de otro artículo. Aquí hay un enlace y un formulario:

[Enlace a google.com](https://www.google.com)

Nombre:

© 2024 Mi Página Web


```
<?xml version="1.0" encoding="UTF-8"?>
<biblioteca>
  <libro>
    <titulo>La vida está en otra parte</titulo>
    <autor>Milan Kundera</autor>
    <fechaPublicacion año="1973"/>
  </libro>
  <libro>
    <titulo>Pantaleón y las visitadoras</titulo>
    <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>
    <fechaPublicacion año="1973"/>
  </libro>
  <libro>
    <titulo>Conversación en la catedral</titulo>
    <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>
    <fechaPublicacion año="1969"/>
  </libro>
</biblioteca>
```



XML es tal cuál como **HTML**; ambos son lenguajes que utilizan etiquetas para construir la estructura de un documento

La principal diferencia es que **XML** suele usarse para **guardar datos estructurados** y compartir entre software (como JSON), y **HTML** se usa para el **contenido en páginas web**

No space allowed

Must have space

No space allowed

```
<p id="myId"></p>
```

Attribute
name

Attribute
value

**En este curso no utilizaremos
XML, sino **XPATH** que es el
lenguaje que extraer
información de XML y HTML**

**Nuestro proceso para hacer
Scraping será traer los HTML
de los sitios web y a través de
XPath extraer la información
que nos importa**

Una expresión de XPATH luce como una instrucción escrita entre '' (comillas simples), que indica qué contenido queremos traer del documento

Ejemplo:

```
'/body/section/article/img[@alt="Un michi en el espacio"]/@src'
```


'/body/section/article/img[@alt="Un michi en el espacio"]/@src'

```
<body>
  <header>
    <h1>Mi Página WEB</h1>
  </header>

  <section>
    <article>
      <h2>Sección 1</h2>
      <p>Este es un párrafo dentro de un artículo. También hay una imagen:</p>
      
      <button>Botoncito</button>
    </article>
  </section>

  <section>
    <article>
      <h2>Sección 2</h2>
      <p>Otro párrafo dentro de otro artículo. Aquí hay un enlace y un formulario:</p>
      <a href="https://www.google.com">Enlace a google.com</a>
      <form>
        <label for="nombre">Nombre:</label>
        <input type="text" id="nombre" name="nombre">
        <button type="submit">Enviar</button>
      </form>
    </article>
  </section>

  <footer>
    <p>&copy; 2024 Mi Página Web</p>
  </footer>
</body>
```

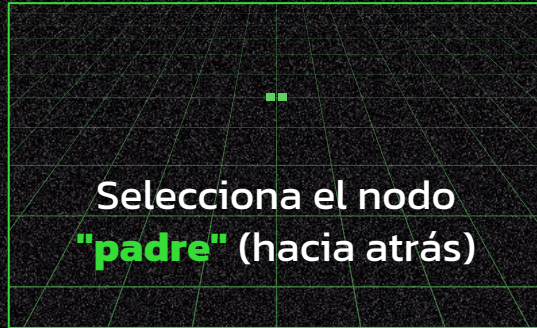
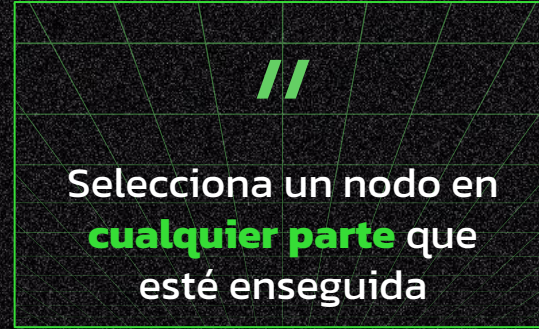
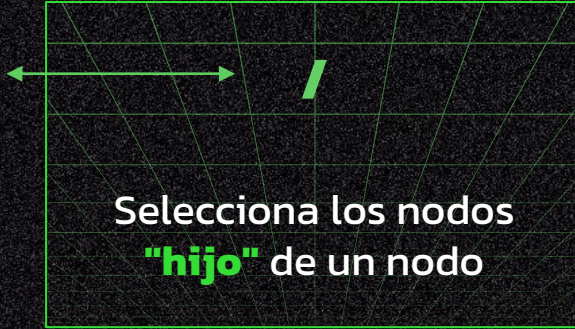
Resultado:

"michi.png"

Expresiones

Navegación entre nodos

También hace referencia a la **"raíz"**, es decir, el **documento completo**




```
<html>
  <head>
    <title>Página Web</title>
  </head>
  <body>
    <div class="block">
      <h2>Título 1</h2>
      <p>Texto número 1</p>
      
      <a href="https://google.com/texto1"></a>
    </div>
    <div class="block">
      <h2>Título 2</h2>
      <p>Texto número 2</p>
      
      <a href="https://google.com/texto2"></a>
    </div>
    <div class="block">
      <h2>Título 3</h2>
      <p>Texto número 3</p>
      
      <a href="https://google.com/texto3"></a>
    </div>
  </body>
</html>
```

EJEMPLOS


```
<html>
  <head>
    <title>Página Web</title>
  </head>
  <body>
    <div class="block">
      <h2>Título 1</h2>
      <p>Texto número 1</p>
      
      <a href="https://google.com/texto1"></a>
    </div>
    <div class="block">
      <h2>Título 2</h2>
      <p>Texto número 2</p>
      
      <a href="https://google.com/texto2"></a>
    </div>
    <div class="block">
      <h2>Título 3</h2>
      <p>Texto número 3</p>
      
      <a href="https://google.com/texto3"></a>
    </div>
  </body>
</html>
```

Selecciona
dentro de /
dentro de **html**
Los nodos **body**

`'/html/body'`


```
<html>
  <head>
    <title>Página Web</title>
  </head>
  <body>
    <div class="block">
      <h2>Título 1</h2>
      <p>Texto número 1</p>
      
      <a href="https://google.com/texto1"></a>
    </div>
    <div class="block">
      <h2>Título 2</h2>
      <p>Texto número 2</p>
      
      <a href="https://google.com/texto2"></a>
    </div>
    <div class="block">
      <h2>Título 3</h2>
      <p>Texto número 3</p>
      
      <a href="https://google.com/texto3"></a>
    </div>
  </body>
</html>
```

Selecciona
dentro de **/**
dentro de **html**
dentro de **body**
dentro de **div**
Los nodos **h2**

'/html/body/div/h2'


```
<html>
  <head>
    <title>Página Web</title>
  </head>
  <body>
    <div class="block">
      <h2>Título 1</h2>
      <p>Texto número 1</p>
      
      <a href="https://google.com/texto1"></a>
    </div>
    <div class="block">
      <h2>Título 2</h2>
      <p>Texto número 2</p>
      
      <a href="https://google.com/texto2"></a>
    </div>
    <div class="block">
      <h2>Título 3</h2>
      <p>Texto número 3</p>
      
      <a href="https://google.com/texto3"></a>
    </div>
  </body>
</html>
```

Selecciona todo
el **/**

Es decir, todo el
documento...

'/'


```
<html>
  <head>
    <title>Página Web</title>
  </head>
  <body>
    <div class="block">
      <h2>Título 1</h2>
      <p>Texto número 1</p>
      
      <a href="https://google.com/texto1"></a>
    </div>
    <div class="block">
      <h2>Título 2</h2>
      <p>Texto número 2</p>
      
      <a href="https://google.com/texto2"></a>
    </div>
    <div class="block">
      <h2>Título 3</h2>
      <p>Texto número 3</p>
      
      <a href="https://google.com/texto3"></a>
    </div>
  </body>
</html>
```

Selecciona
todos los nodos
h2
dentro del /

'//h2'

**RECUERDA QUE EL `"//"` ES
COMO SALTAR A
CUALQUIER NODO, SIN
HACER EL RECORRIDO
COMPLETO**

Selecciona
todos los nodos
dentro del `h2`

`'//h2'`


```
<html>
  <head>
    <title>Página Web</title>
  </head>
  <body>
    <div class="block">
      <h2>Título 1</h2>
      <p>Texto numero 1</p>
      
      <a href="https://google.com/texto1"></a>
    </div>
    <div class="block">
      <h2>Título 2</h2>
      <p>Texto número 2</p>
      
      <a href="https://google.com/texto2"></a>
    </div>
    <div class="block">
      <h2>Título 3</h2>
      <p>Texto numero 3</p>
      
      <a href="https://google.com/texto3"></a>
    </div>
  </body>
</html>
```

Selecciona todo
nodo **h2** en este
caso de **HTML**

'//h2'

'/html/body/div/h2'

Seleccionan exactamente lo mismo
aunque no hacen lo mismo


```
<html>
  <head>
    <title>Página Web</title>
  </head>
  <body>
    <div class="block">
      <h2>Título 1</h2>
      <p>Texto número 1</p>
      
      <a href="https://google.com/texto1"></a>
    </div>
    <div class="block">
      <h2>Título 2</h2>
      <p>Texto número 2</p>
      
      <a href="https://google.com/texto2"></a>
    </div>
    <div class="block">
      <h2>Título 3</h2>
      <p>Texto número 3</p>
      
      <a href="https://google.com/texto3"></a>
    </div>
  </body>
</html>
```

Selecciona
todos los nodos
img
dentro del /

'//img'


```
<html>
  <head>
    <title>Página Web</title>
  </head>
  <body>
    <div class="block">
      <h2>Título 1</h2>
      <p>Texto número 1</p>
      
      <a href="https://google.com/texto1"></a>
    </div>
    <div class="block">
      <h2>Título 2</h2>
      <p>Texto número 2</p>
      
      <a href="https://google.com/texto2"></a>
    </div>
    <div class="block">
      <h2>Título 3</h2>
      <p>Texto número 3</p>
      
      <a href="https://google.com/texto3"></a>
    </div>
  </body>
</html>
```

Selecciona de
todo **img**
dentro del **/**
El nodo **padre**

'//img/..'

**ES IMPORTANTE
ACLARAR QUE
LAS XPATH
REGRESA SUS
RESPUESTAS EN
UNA LISTA DE
ELEMENTOS**

```
<html>
  <head>
    <title>Página Web</title>
  </head>
  <body>
    <div class="block">
      <h2>Título 1</h2>
      <p>Texto número 1</p>
      
      <a href="https://google.com/texto1"></a>
    </div>
    <div class="block">
      <h2>Título 2</h2>
      <p>Texto número 2</p>
      
      <a href="https://google.com/texto2"></a>
    </div>
    <div class="block">
      <h2>Título 3</h2>
      <p>Texto número 3</p>
      
      <a href="https://google.com/texto3"></a>
    </div>
  </body>
</html>
```

Selecciona de
todo **img**
dentro del **/**
El nodo **padre**

`'//img/..'`


```

<html>
  <head>
    <title>Página Web</title>
  </head>
  <body>
    <div class="block">
      <h2>Título 1</h2>
      <p>Texto número 1</p>
      
      <a href="https://google.com/texto1"></a>
    </div>
    <div class="block">
      <h2>Título 2</h2>
      <p>Texto número 2</p>
      
      <a href="https://google.com/texto2"></a>
    </div>
    <div class="block">
      <h2>Título 3</h2>
      <p>Texto número 3</p>
      
      <a href="https://google.com/texto3"></a>
    </div>
  </body>
</html>

```

Selecciona de
todo **img**
dentro del **/**
del nodo **padre**
los nodos **img**

'//img/ ../img'


```

<html>
  <head>
    <title>Página Web</title>
  </head>
  <body>
    <div class="block">
      <h2>Título 1</h2>
      <p>Texto número 1</p>
      
      <a href="https://google.com/texto1"></a>
    </div>
    <div class="block">
      <h2>Título 2</h2>
      <p>Texto número 2</p>
      
      <a href="https://google.com/texto2"></a>
    </div>
    <div class="block">
      <h2>Título 3</h2>
      <p>Texto número 3</p>
      
      <a href="https://google.com/texto3"></a>
    </div>
  </body>
</html>

```

Selecciona
todos los nodos
img

'//img'

'//img/../../img/../../img/../../img'

Hacen exactamente lo
mismo


```

<html>
  <head>
    <title>Página Web</title>
  </head>
  <body>
    <div class="block">
      <h2>Título 1</h2>
      <p>Texto número 1</p>
      
      <a href="https://google.com/texto1"></a>
    </div>
    <div class="block">
      <h2>Título 2</h2>
      <p>Texto número 2</p>
      
      <a href="https://google.com/texto2"></a>
    </div>
    <div class="block">
      <h2>Título 3</h2>
      <p>Texto número 3</p>
      
      <a href="https://google.com/texto3"></a>
    </div>
  </body>
</html>

```

Selecciona
todos los nodos
img

'//img'

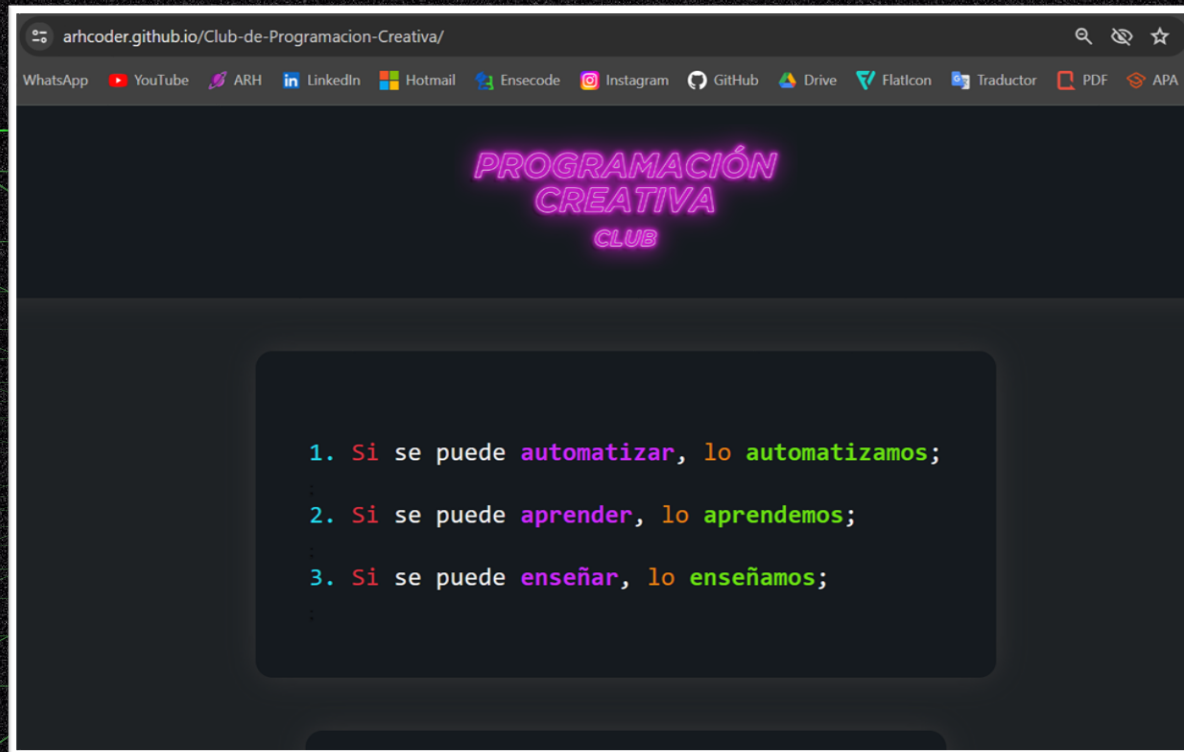
'//img/.'

'//img/./././././.'

Hacen exactamente lo mismo



PÓNLO A PRUEBA



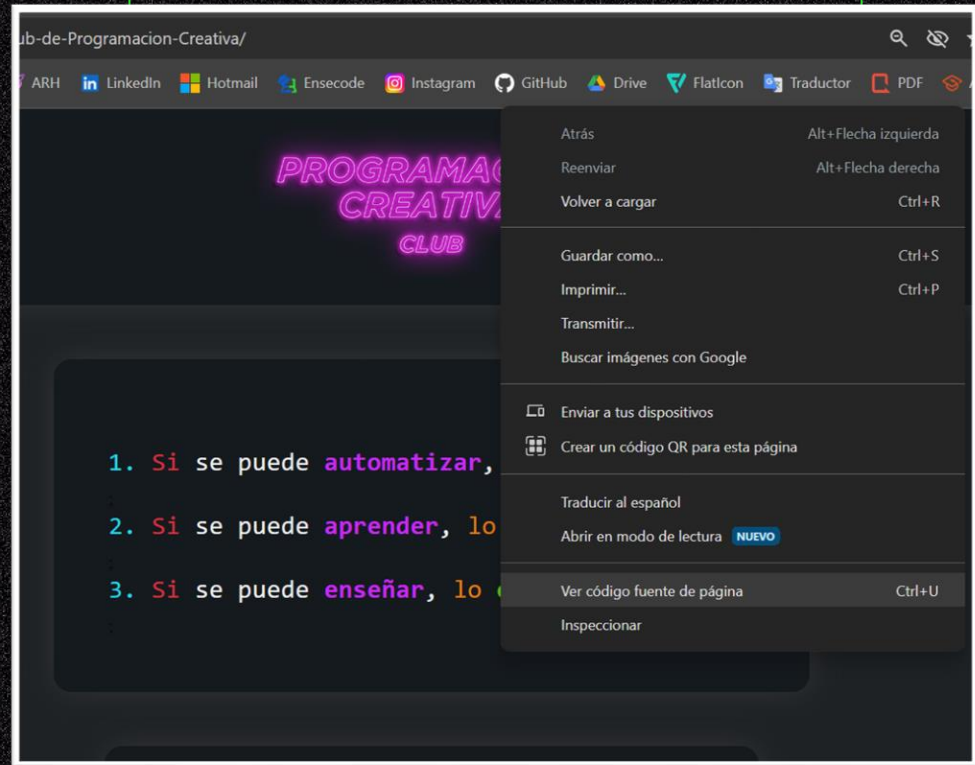
Entra a: <https://arhcoder.github.io/Club-de-Programacion-Creativa/>

IMPORTANTE

PARA VER EL CÓDIGO FUENTE HTML:

1. Abre el la página web
2. Click derecho en la página
3. "Ver código fuente de página"

También funciona con Ctrl + U



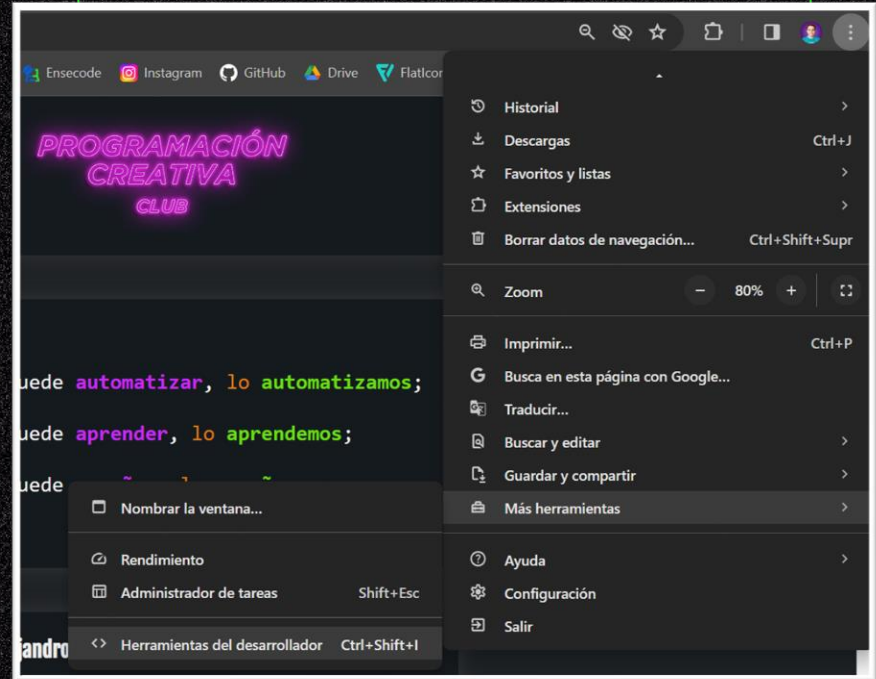
IMPORTANTE

PARA PRACTICAR XPATH:

1. Abre el sitio web
2. En la barra de opciones del navegador >> Más Herramientas >> Herramientas del Desarrollador (Inspector de Elementos)
3. Abre la pestaña "**Consola**"
4. Para ejecutar una consulta XPATH:

`$x('expresion/xpath/aqui')`

También funciona con Ctrl + Shift + I



The image features a black background with a green wireframe grid. The grid consists of horizontal and vertical lines, with additional diagonal lines crossing the entire area. In the center of this grid, the word "Ejercicios" is written in a bold, green, sans-serif font.

Ejercicios

Ejercicios

1. Extraer todos los **img** del sitio web:

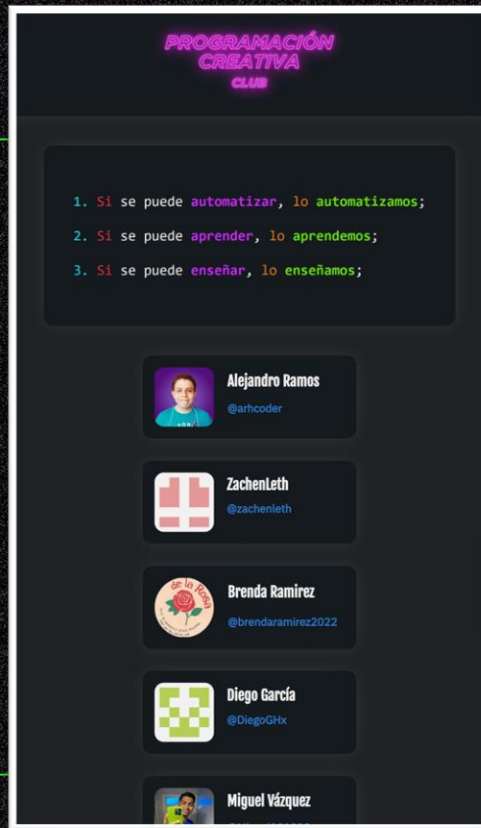
```
$x('')
```

2. Extraer todos los **textos** de las 3 reglas del Club:

```
$x('')
```

3. Extraer los **img** de foto de usuario:

```
$x('')
```



<https://arhcoder.github.io/Club-de-Programacion-Creativa/>

Respuestas

1. Extraer todos los **img** del sitio web:

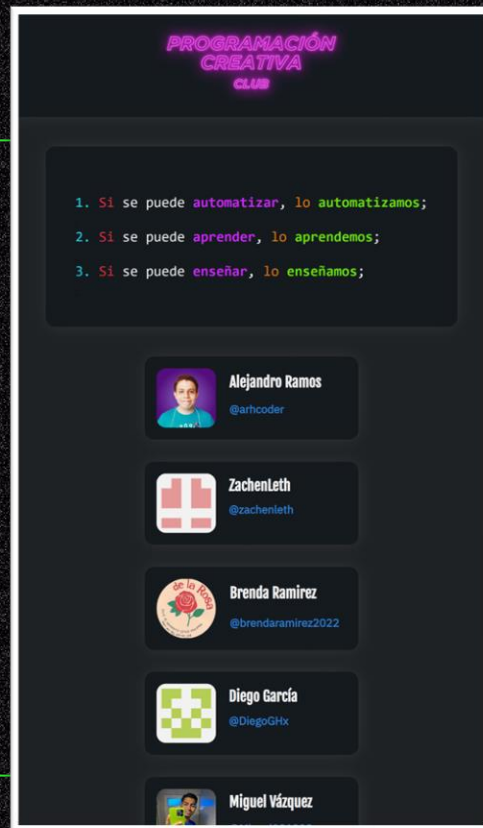
```
$x('//img')
```

2. Extraer todos los **textos** de las 3 reglas del Club:

```
$x('//span')
```

3. Extraer los **img** de foto de usuario:

```
$x('/html/body//div/img')
```



<https://arhcoder.github.io/Club-de-Programacion-Creativa/>

The image features a black background with a green wireframe grid. The grid consists of horizontal and vertical lines, with additional diagonal lines forming a perspective effect that converges towards the center. In the center of this grid, the word "PREDICADOS" is written in a bold, green, sans-serif font.

PREDICADOS

Predicados

Los predicados son **"condicionales"**.
Son expresiones lógicas con las que se **filtra** lo que queremos y lo que no

Se colocan siempre **entre []**

Posibles predicados de selección:

- **Selección por atributos**
- **Selección por posición**
- **Selección por contenido del texto**
- **Selección por ancestros**

```
$x('//span[@class!="text"]') // todos  
los span que tenga una clase distinta  
de "text"
```

```
$x('/html/body/div/div[position()=1]')  
// trae los elementos en la posición 1
```

```
$x('/html/body/div/div[position()>1]')  
// trae todos los elementos que se  
encuentran después de la posición 1
```

```
$x('//span[@class="text" and  
@class="tag-item"]') // trae los  
elementos que tengan como clase a  
"text" Y a "tag-item"
```

```
$x('//span[@class="text" or  
@class="tag-item"]') // trae los  
elementos que tengan como clase a  
"text" O a "tag-item"
```

```
$x('//span[not(@class)]') // trae  
todos los span que NO tengan una clase
```



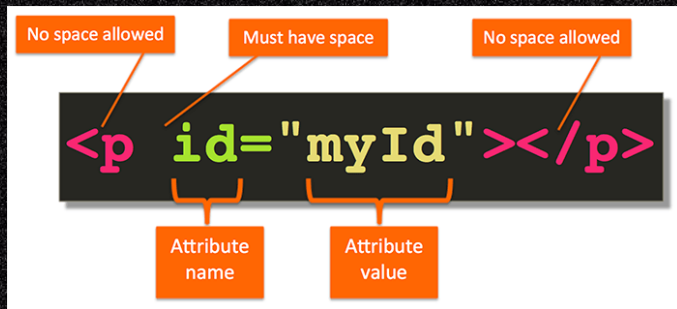

Selección por Atributos

Selección por atributos

Si queremos seleccionar nodos **que tengan o no un atributo**.

Es **muy muy muy útil**, sobre todo porque en Desarrollo Web a menudo los nodos se suelen diferenciar del resto por dos atributos comunes: **"class"**, **"id"**.

Para referenciar un **nombre de atributo** en XPATH se utiliza **@**.



Selección por atributos

Ejemplos:

```
$x('//div[@class = "calendario"]')
```

Todos los nodos **div** que tienen el atributo **class = "calendario"**

```
$x('//span[@class != "texto"]')
```

Todos los nodos **span** que tienen el atributo **class** diferente de **"texto"**

```
$x('//button[@id = "submit"]')
```

Todos los **button** con el atributo **id = "submit"**

```
$x('//img[@width = "800px"]')
```

Todas las imágenes con un ancho de 800 px



**NOTA
IMPORTANTE**

Operadores Lógicos

Dentro de los predicados "[]" podemos colocar más de una condicional, anidándolas con **operadores lógicos**:

- **and**
- **or**
- **not**

O utilizar otros operadores como:

- **>**
- **<**
- **>=**
- **<=**

Operadores Lógicos

Ejemplos:

```
$x('//div[@class = "calendario" or  
@class = "reloj"]')
```

Todos los nodos **div** que tienen el atributo **class = "calendario"** o **class = "reloj"**

```
$x('//button[@id = "submit" and  
@class = "login"]')
```

Todos los **button** con el atributo **id = "submit"** y además el atributo **class = "login"**

Operadores Lógicos

Ejemplos:

**RECUERDA QUE SI SE TIENEN VARIAS
CONDICIONALES PARA UN NODO, SE**

**COLOCAN TODAS DENTRO DEL MISMO
CORCHETE. EJEMPLO:**

```
$x('//img[@class="profesor" or @class="alumno"]')
```

```
$  
@class = "login"]')
```

```
uto  
y además el atributo  
class = "login"
```


Operadores Lógicos

Ejemplos:

Para el caso del operador "**not**", hará falta siempre utilizar **paréntesis** para agrupar:

```
$x('//div[not(@class)]')
```

Todos los **div** que NO tengan el atributo **class**

ES ÚTIL SABER QUE EN XPATH, LOS PARÉNTESIS "(")" SERVIRÁN PARA AGRUPAR, SOBRE TODO CUANDO QUERAMOS APLICAR ALGUNOS PREDICADOS A EXPRESIONES COMPLETAS; TAL CUÁL COMO EN MATEMÁTICAS...

The image features a black background with a green wireframe grid. The grid consists of horizontal and vertical lines, with additional diagonal lines crossing the grid at an angle. In the center of the grid, the word "Ejercicios" is written in a bold, green, sans-serif font.

Ejercicios

Ejercicios

1. Extraer todos los **textos** morados:

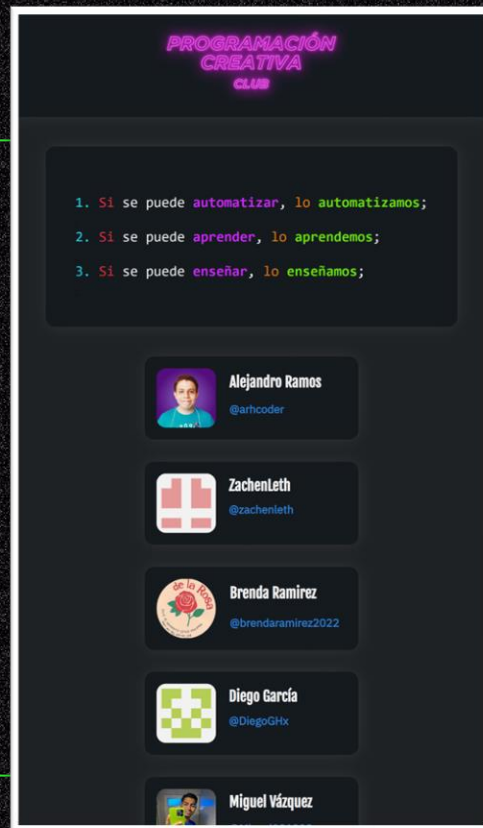
`$x('')`

2. Extraer todos los **bloques** con la info y foto de los miembros:

`$x('')`

3. Extraer los **img** de foto de usuario:

`$x('')`



<https://arhcoder.github.io/Club-de-Programacion-Creativa/>

Respuestas

1. Extraer todos los **textos** morados:

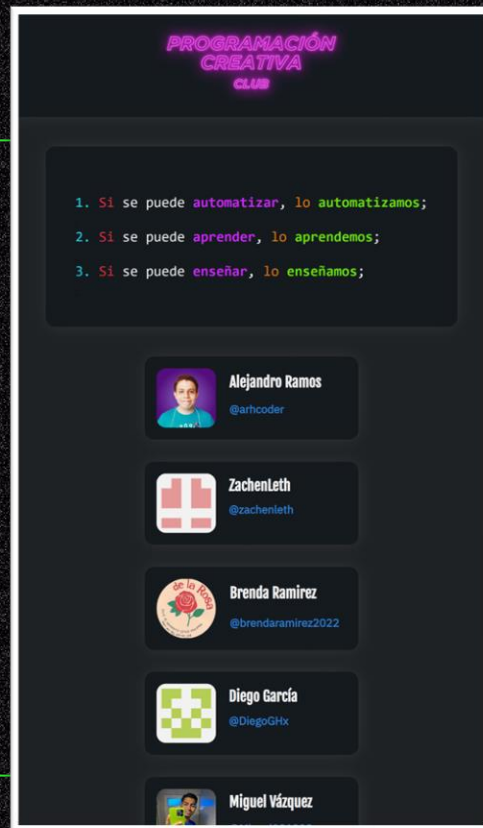
```
$x('//span[@class="purple-text"]')
```

2. Extraer todos los **bloques** con la info y foto de los miembros:

```
$x('//div[@class="member"]')
```

3. Extraer los **img** de foto de usuario:

```
$x('//img[@class="avatar"]')
```



<https://arhcoder.github.io/Club-de-Programacion-Creativa/>



Selección por posición

Selección por posición

Recordemos que una consulta de **XPATH** **regresa siempre una lista de elementos (incluso si está vacía o si sólo contiene un ítem)**

Por ejemplo:

```
$x('//div[@class="member"]')
```

Regresa la lista de **div**'s con la **clase "member"**

¿Qué tal si quisiéramos que nos regresara sólo ciertos miembros?

- **[n]**
- **[position() > n]**
- **[last()]**

Selección por posición [n]

Ejemplos:

```
$x('(//div[@class = "member"])[1]')
```

De los **div** con **class = "member"**
selecciona el **PRIMERO [1]**

```
$x('(//div[@class = "member"])[4]')
```

De los **div** con **class = "member"**
selecciona el **CUARTO [4]**

Selección por posición [n]

NOTA CÓMO SE ESTÁN UTILIZANDO

Ejemplos:

PARÉNTESIS PARA AGRUPAR, Y QUE

LOS "[n]" AFECTEN A TODO EL

RESULTADO DE LA CONSULTA

```
$x('(//div[@class="member"])[1]')
```

selecciona el PRIMERO [1]

```
$x('(//div[@class="member"])[4]')
```

```
$x('(//div[@class="member"])[4]')
```

selecciona el CUARTO [4]

EN ESTE CASO **NO HACEN FALTA LOS PARÉNTESIS;**
PERO ES BUENO ACOSTUMBRARSE A PONERLOS,
PUES SI TUVIÉRAMOS UNA **EXPRESIÓN MÁS LARGA, SÍ**
HARÍAN FALTA...

```
$x('←//div[@class="member"]→[4]')
```

```
$x(' (/html/body/div//img)[4]')
```


Selección por posición [position()]

Ejemplos:

```
$x('//div[position() > 4]')
```

De los **div**, selecciona todos
después del cuarto

```
$x('//div[@class="member"]  
[position() >= 4]')
```

De los **div** con **class = "member"**
Selecciona todos a **partir del cuarto**

Selección por posición [position()]

Ejemplos:

```
$x('//a[last()])')
```

De los **a**, selecciona el **último**

```
$x('( /html/body/div//img)[last()-1]')
```

De todos los **img** después de
/html/body/div/
Selecciona el **penúltimo**

The image features a black background with a green wireframe grid. The grid consists of horizontal and vertical lines, with additional diagonal lines crossing the grid at an angle. In the center of the grid, the word "Ejercicios" is written in a bold, green, sans-serif font.

Ejercicios

EJERCICIOS

Describe lo que hacen las siguientes expresiones:

```
$x(' (/html/body/div[@class="content"]  
//img)[last()-2]')
```

Descripción...

```
$x('/html/body/div[@class="content"]/  
/img[last()-2]')
```

Descripción...

EJERCICIOS

Describe lo que hacen las siguientes expresiones:

```
$x(' (/html/body/div[@class="content"]  
//img)[last()-2]')
```

Después de los **div** en **/html/body** con **clase "content"**, selecciona la antepenúltima **img**

```
$x('/html/body/div[@class="content"]/  
/img[last()-2]')
```

Lo mismo que la de arriba, pero al **faltar paréntesis** para agrupar, no selecciona nada...

Quotes to Scrape

"The world as we have created it is a process of our thinking. It cannot be changed without changing our thinking."

by [Albert Einstein](#) (about)

Tags: [change](#) [deep-thoughts](#) [thinking](#) [world](#)

"It is our choices, Harry, that show what we truly are, far more than our abilities."

by [J.K. Rowling](#) (about)

Tags: [abilities](#) [choices](#)

"There are only two ways to live your life. One is as though nothing is a miracle. The other is as though everything is a miracle."

by [Albert Einstein](#) (about)

Tags: [inspirational](#) [life](#) [live](#) [miracle](#) [miracles](#)

Para practicar Scraping, entra en: <https://quotes.toscrape.com/>



Continuará...

En 02 - B

Club de Programación Creativa En YouTube



Alejandro Ramos
@arhcoder